08 | 怎样搞定一场用户可用性测试？

2021-01-25 炒炒

体验设计案例课 进入课程 >

讲述：炒炒
时长 18:56 大小 17.35M

你好，我是炒炒。

在工作中，你是不是经常会听到需求方说下面几句话：

> 我感觉你这个方案也不是用户想要的吧？
>
> 你这样设计会不会让用户觉得很麻烦呀？
>
> 你不觉得这样做效率很低吗？

听到这些，你肯定会想“你又不是用户，你凭什么就确定这不是用户想要的？ “。

为什么我们会有这种分歧？其实，分歧产生的根本原因就是我们对用户视角理解的不一样。那么，我们和用户所思所想之间真的有差距吗？到底我们设计的方案是不是用户想要的方案？

问题的答案，设计师和需求方说了都不算，只有用户说了才算。所以，为了能够进一步统一双方的用户视角，我们就可以通过一次用户可用性测试来解决这个问题。

什么是用户可用性测试？

做过用户体验设计的同学应该比较了解用户可用性测试，用户可用性测试是一种测试方法。

如果你不了解也没关系，我们单从字面上来看，也可以理解什么是用户可用性测试。就是**让用户做主角，我们再去通过观察用户的行为来完成我们的测试任务**，这个过程就是用户可用性测试。

这种方法可以让我们站在用户的视角上思考，去发现我们产品设计中存在的问题。而且，这种方法还解决了我们和需求方之间的分歧，也就是到底该听谁的？那自然是听用户的。

现在你知道用户可用性测试是什么了，也许你觉得这件事很简单，没有那么复杂。

的确，用户可用性测试是没有那么复杂，但是在这个过程中，我们经常会弄混一件事情，也是我想跟你强调的一点：可用性测试的目标到底是什么。

我用个例子来说明一下。我之前做过一个 B 端产品，产品上线后，我们邀请了目标用户，来做用户可用性测试。不过，当时用户突然操作不下去，这个时候，主持人就开始进行产品讲解，用户不断地提出问题，主持人很耐心地回答和解说。

你发现问题了吗？问题就在于，这本是一场用户可用性测试，却演变成了一个小型产品说明会。

不仅用户一脸懵，而且我们也没获取到真实有效的产品体验意见和感受，双方对此都很不满意，花了时间和钱并没有办成事情。究其根本，就是因为我们的目标不清晰。

我们的核心目标是用户的感受，而不是我们强加给他们的感受，我们的任务是观察用户的行为，然后引导用户说出他的感受，我们只要真实记录下来就可以了。

接下来，我们就聊聊如何搞定一场用户可用性测试。

在具体实操中，我们通常把一场完整的用户可用性测试分为 5 个步骤：**设定典型任务、邀请目标用户、组织测试、度量分析、输出报告。**

在接下来的内容里，我会带你逐一了解每个步骤里面常见的一些坑以及如何避免。

让我们开始吧！

第一步，设定典型任务

在描述典型任务的时候，一定要做到清晰简练，避免过于简单或者过于复杂的任务描述。

假如我们收到了一个这样的任务描述“请完成理财产品 A 的购买”。你看到这个任务后是不是觉得很简单，但是却不知道如何下手？

如果把这个任务放进我们企业银行理财购买的场景中来看，你会发现有较多的漏洞。

在我们要测的产品里，有好几条路径都可以完成理财产品 A 的购买，每个用户操作习惯不同，会根据自己的操作习惯去购买理财产品 A。这个任务并没有说清楚任务开始于哪个页面，任务结束在哪个节点。这样，我们无法准确计算任务完成所需要的时长。

如果我们把这个任务描述优化一下，改成下面这样呢？“请你通过某产品首页进入理财超市，并在理财超市找到理财产品 A，购买 10 元理财产品 A，支付成功视为任务结束”。

你看，优化后的描述是不是就清晰地说明了任务开始和结束的标志，路径也说的很明确，这样的任务才能算是合格的任务。

当然，除了设定好典型任务，在任务收集过程中，常常会出现一种情况，各个业务部门因自己负责的领域不同，大家想测的任务也不同。怎么解决呢？我们要把所有任务中的相似任务进行整合，再根据需求优先级和功能的使用频率整理出少量的，但是典型的任务。

在我们一次的测试中，我们还遇到了任务设定过于复杂的情况。面对过于复杂的任务，我们要学会结合目标对任务进行精简，不要给用户增加不必要的操作负担。

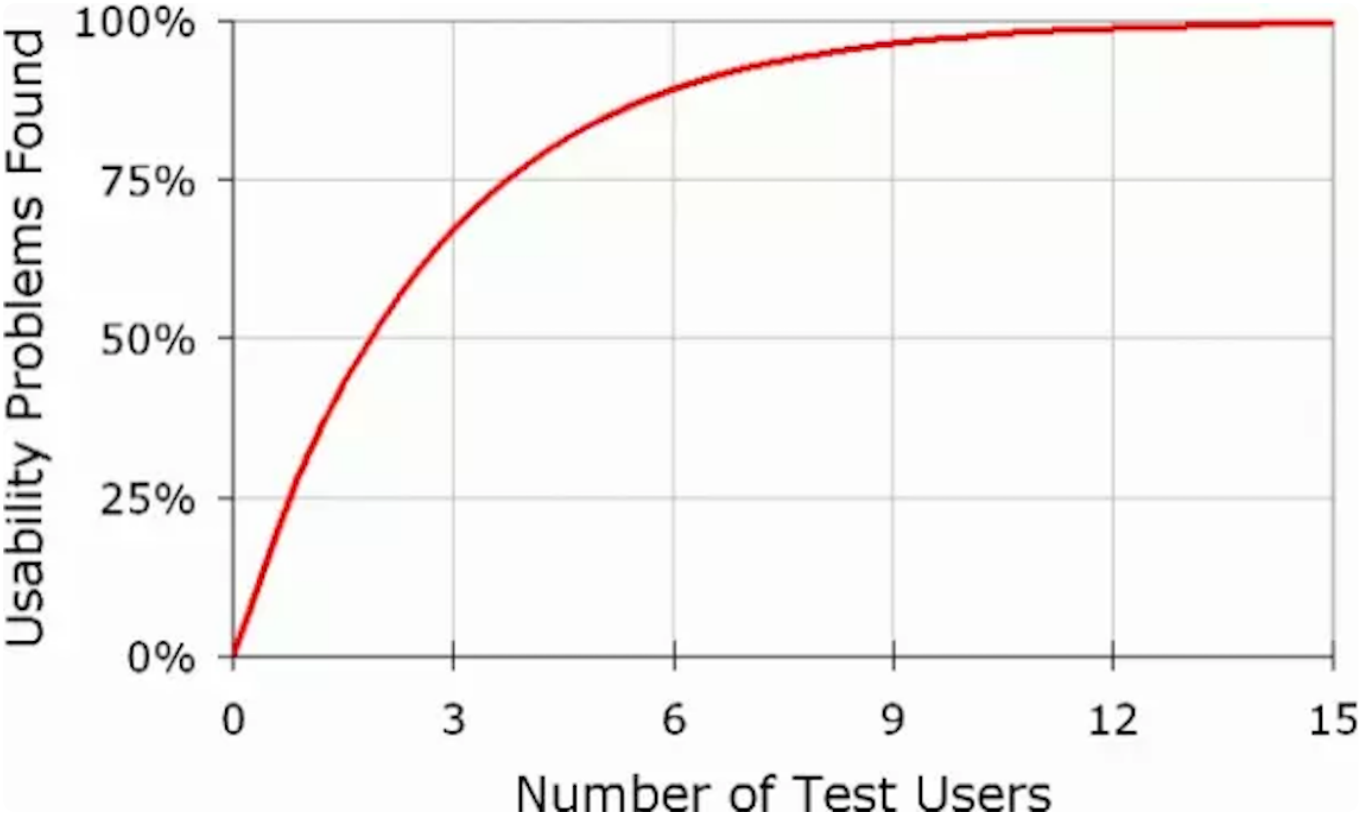
最后，我们整理出 10 个备选任务之后，把它们拿到了项目例会上，通过投票的方式确定了最终的 5 个典型任务。当然，在投票前，我们要解释清楚合并的逻辑以及为什么是这 10 条的原因。

第二步，邀请代表性用户

在确定了典型任务后，我们就可以邀请目标用户了。

很多人都觉得邀请的用户越多越好，这样样本量大，能得出更准确的结论。真的是这样吗？不是的，用户越多，意味着时间成本和精力成本也越高。

在这里，不得不先说一下 Nielsen 提出的一个法则：有 5 人参加的用户可用性测试，即可发现 85% 的产品可用性问题，而且通常最严重的问题都是前几名用户发现的，随着用户数量的增多，发现的问题逐渐减少，被发现问题的数量与测试用户数量的关系如下图所示。



结合 Nielsen 的研究结论，我们开始商定了邀请 8 名目标用户，但其他协同部门认为 8 个人太少了，覆盖的不够全面，他们想要邀请更多的目标用户参与进来。最终在协作方的要

求下，我们又增加了 4 个用户，这样总数就变成了 12 名用户。

那这 12 名用户我们是如何挑选出来的呢？答案是**限定条件下的随机**。

我们会根据我们的用户画像，随机挑选 12 名我们理财的潜在客户——曾经浏览过我们的理财产品，或者曾经购买过同类产品，且是一个月出现过的活跃用户，并在我们所在的当地。

挑选出用户，我们会提前跟目标用户沟通，询问用户是否愿意加入我们的测试计划中，我们会提供小礼物作为感谢，获得同意的用户才会收到正式的邀请函。

在后续的测试中，我们以每天 4 个用户的进度开展测试。在第一天结束后，经过去重汇总，我们得到了 76 个问题，第二天结束后，去重汇总后的问题数量变成了 86 个，到了第三天结束后，去重汇总后的问题数量变成了 92 个。

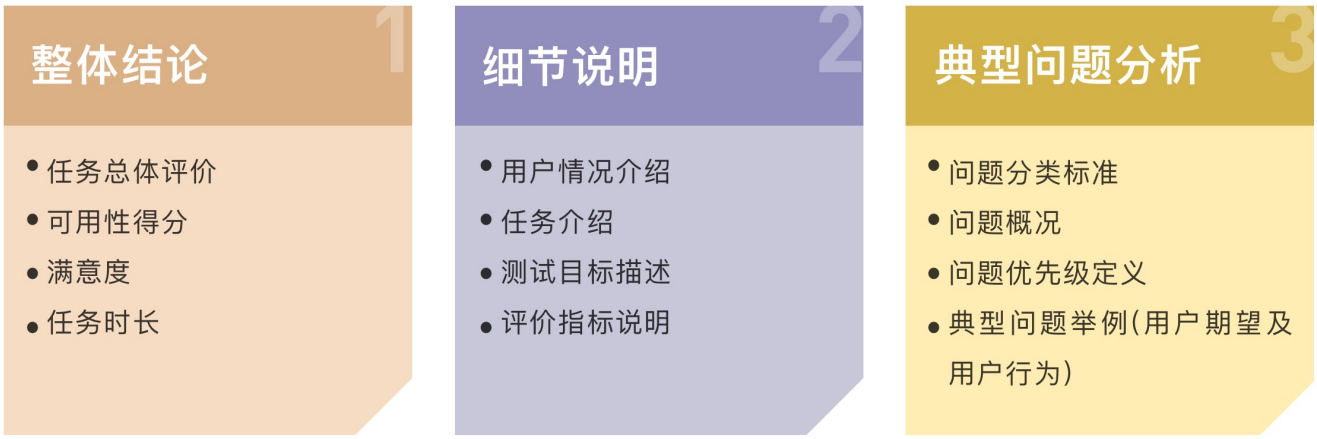
通过我们实际的操作数据可以发现，第二天和第三天的用户遇到的新增问题数量分别为 10 个和 6 个，并没有第一天的 4 个用户发现的问题数量多。**新发现问题的数量会逐步呈现递减**。

这个结果和 Nielsen 的研究非常接近。所以，我们在确定目标用户数量的时候，可以直接参考 Nielsen 研究结论，邀请 5-8 名目标用户就足够了。邀请的用户越多，我们投入的精力越大，但发现问题的数量并不是成对应比例的增加，尤其是在时间和精力都比较有限的情况下。

第三步，组织测试

确定了我们的典型任务，也基本敲定了目标用户，然后我们就可以动手准备组织测试了。

是不是摩拳擦掌，跃跃欲试？先别着急，在正式测试开始前，我们可以先找身边的同事做一次预测试，通过测试，我们可以找到一些容易忽略的细节问题。



我们以“账户余额不足时，请转账 10 元到理财账户”为任务，做一次预测试，发现由于某个账号权限不能使用而导致该任务未能完成。事后找了对应的测试沟通才发现，原来有另外一个同事刚刚用这个账号测了另外一个流程，把对应的权限关闭了，所以导致了预测试中的任务失败。

有了这个教训，我们又检查了一下准备的所有账号，并把账号设置成了用户可用性测试专用，这样就确保在接下来的测试过程中，不会因为账号权限的问题影响用户可用性测试的正常进行。

接下来，在用户可用性测试的过程中，**我们要尽量鼓励用户使用发声法来参与测试**。引导用户在操作任务的过程中，将想法即时地表达出来。通常情况下包含了他们正在做什么、想要做什么、预期是什么、这样做的原因是什么等等。

发声法有利于我们收集到更多用户对产品细节的想法，判断逻辑，决策要素等，但用户不太习惯采用发声法参与测试。这时候，需要我们对用户的表情以及操作方式进行敏锐地觉察与记录。

这里有两个技巧能够提升一些用户采用发声法的意愿。

第一个技巧是在真实开始前预留 10 分钟左右的闲聊时间，拉近自己和用户之间的距离，努力让双方变得更熟悉，减少用户的压力。比如说有一个在企业做财务的用户，他过来后非常谨慎，问我们说“会不会造成个人信息泄漏” “会不会泄露他公司的财务数据”等等。

我们会开门见山直接说，我们只是对我们产品的使用情况进行一个测试，看看像您这样的天使用户对我们产品的看法，好不好用，不好用我们改，未来给您提供更好的服务。

听完这个信息，受访者会明显放松下来。这个时候，聊聊生意好做不好做，行业趣闻，一下子就可以打开用户可用性测试的局面。

二是给用户演示发声法的测试方法，让用户直观地感受到原来也可以这样参加测试。如果用户实在不愿意采用发声法，也不用强制要求，加强在测试过程中对用户的一些微表情的观察就好。

在整个过程中，主持人和观察员要时刻保持**客观记录**原则，不能以主观的意愿影响用户的操作。

我们遇到过一个用户，他可能比较紧张，在操作的时候不停地问我，他操作的对不对。他每次这样发问的时候，我们就会鼓励他说“没问题的，你做的都对”“没关系，你只要按照你想的操作来就好了，这个没有标准的要求”“这个没有对错，你开心就好”。

总之就是不能以我的主观意愿告诉他做的对不对，我们只需要多鼓励，尽量帮他缓解紧张情绪，真实客观记录就好。

这里还有个提示，**就是当天收集到的问题和用户资料**，当天做好整理。这样做的好处是能够帮我们在整理问题的时候回忆起更多的细节。问题整理出来后，对于今天的可用性测试结论会有一个初步的掌握，复盘操作中的突发状况和解决方案，在接下来的测试中尽力避免，有效引导 。

同时，也能对一些已掌握的问题进行重点观察，并验证已经得出的一些推论。

第四步，度量分析

在测试任务完成后，就可以开始对本次的测试结果进行度量分析了。也就是说，把大量的用户可用性测试原始资料转换成能对产品优化形成指导意见的材料。

通常来说，我们会从两个角度来进行度量，一个是可用性度量，一个是问题度量。

可用性度量指的是从任务流程的完成率、效率、满意度维度来衡量任务的可用性和必要性；

问题度量指的是问题的分类和优先级。

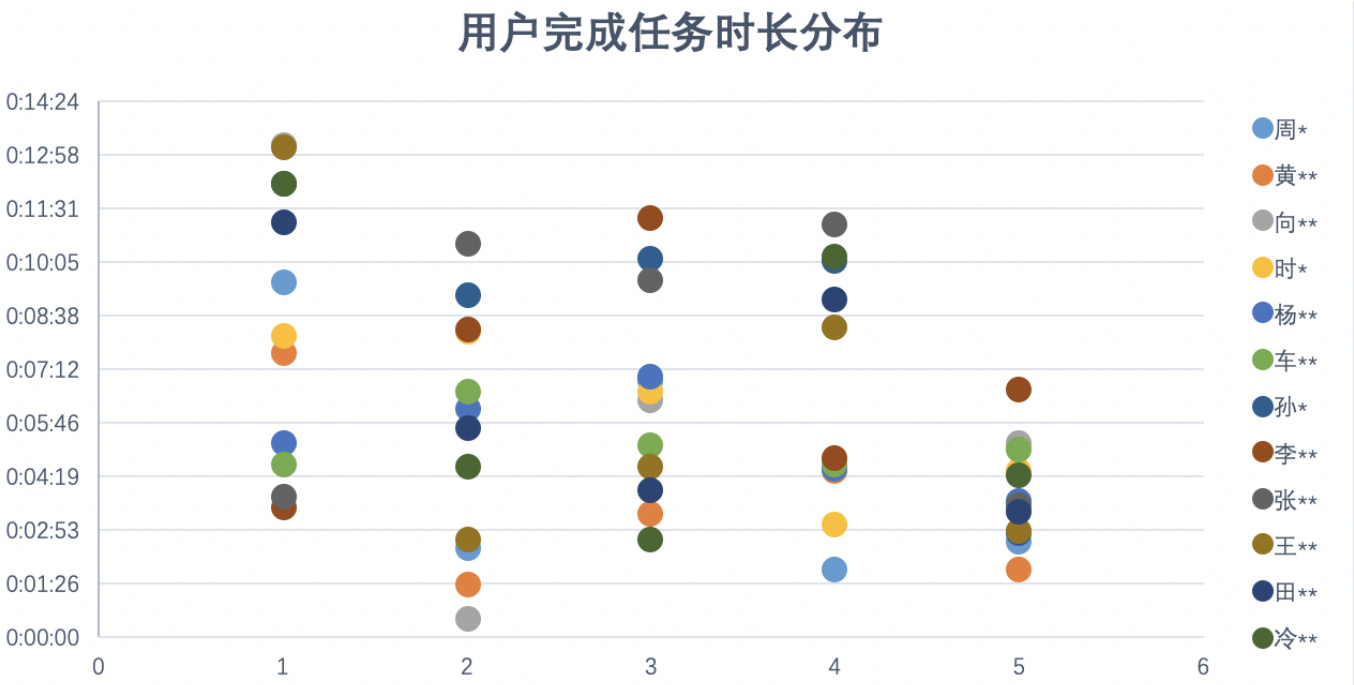
在可用性度量层面，常规的完成率和满意度的计算方法我就不在这里讲了，想跟你分享一下我们本次测试采用的一个特殊的效率评估指标。通常的效率评估会看平均时长，但这次我们没有看平均时长，看的却是时长分布。

为什么看时长分布却不看平均时长呢？平均时长一般用来看 c 端通用型任务，例如添加一个好友或者是完成一次支付。平均时长越短，从操作层面上看，体验越好。

本次的典型任务对应的业务流程都相对固定，不能进行流程的删减或改变。这种情况下看时长分布反而更能看出**用户对这些固定流程的理解是否一致。**

比如说我们这次的时长分布结果图（如下），你看在任务二中，每个用户的完成时长分布非常分散，而且跨度比较大，这就说明不同用户之间对该任务理解的一致性较差，这也从侧面反映了该任务相关的流程页面有比较大的改进空间；

再看一下任务五，每个用户完成的时长分布相对集中，这就说明不同用户对该任务理解的一致性较好，流程和页面的设计也相对合理。



在问题度量层面，常规情况下给问题分类和定义优先级是必须要做的两件事情。

这里我想跟你分享一点容易被忽略的地方：**有效区分真问题和假问题。**

如果同一问题多次在不同的用户之间出现，那这个问题就可以定义为一个真问题；如果一个问题仅是在某一个用户的操作行为中偶然出现，并且该用户也说不清楚对应的行为逻辑，那这个问题可以定义为一个假问题。但如果这种偶然出现的问题，我们发现用户的行为决策、思考过程是符合逻辑的，那这个问题就需要定义为真问题。

举个例子，在转账完成后，用户需要通知复核人员去复核。用户张三和李四都没点击页面上的通知复核按钮。通过进一步询问后得知，张三说他更习惯直接打电话通知复核人员；李四说他没有找到通知复核的按钮，所以才选择打电话通知复核人员。

从询问的结果来看，张三是因为个人习惯没有点击通知复核按钮，李四是因为找不到通知复核按钮才没有点击，张三的场景就是一个假问题，李四的场景就是一个真问题。

第五步，输出报告

我们做可用性测试报告的核心目的，是希望产品能通过迭代优化变得更好。

如何把结论更好地输出给相关的协作方？不是越多越好，我们对结论也需要提炼总结，对产品体验的影响度需要有标识，并且要给出建议的解决方案。

本次我们输出了两个文件：用户可用性测试报告和问题清单。

用户可用性测试报告用来对本次测试做一个概括性的总结，方便跟相关领导和同事可以快速地对齐产品问题，并对用户可用性测试的相关情况有个大致的了解。

用户可用性测试报告的编写逻辑，推荐你用**整体结论 - 详情介绍 - 典型问题分析**，这种逻辑在汇报的时候特别高效。因为领导通常没有太多的时间听你平铺直叙地讲完所有的测试细节，所以先把结论给出来，支撑结论的论点论据后面再讲。

我们再分别看一下每个部分的细节内容。

整体结论。包含了对任务的可用性得分汇总、任务时长分布以及每个任务的整体评价。概括性的结论可以帮助其他人对产品的本次测试状况有个快速的直观了解；

用户可用性测试过程及资料的详细说明。包含了用户的情况、任务描述、本次测试的目标、评价用的相关指标说明等。这样想了解测试背景和详情的人可以通过这部分内容来了解。

典型问题的分析。包含了问题分类方式、整体问题数量、典型问题举例以及该问题的用户行为和用户期望等。这部分内容方便其他人对用户遇到的问题有了解，并能够对典型问题保持关注。

可用性测试报告更适合整体汇报的场景，那么针对下一步的问题优化，我们该怎样做会有更好的效果呢？其实，我们通常采用 Excel 格式制作问题汇总跟踪表，简单又高效。

问题汇总表一般会包含以下要素：序号、问题描述、建议方案、严重程度、出现频率、问题优先级、问题类型、跟进人、进度。通过这些要素，我们就可以直观地看到每个问题的

紧急程度以及后续的进度，这样才能督促问题的优化落地。

炒炒总结

今天，我们探讨了一个话题，如何搞定一场用户可用性测试。虽然，这一讲的内容看起来有点多，但是一定要记住一点，就是可用性测试的目的一定是为了获取最真实的用户意见。

可用性测试包含了 5 个步骤：设定典型任务 + 邀请目标用户 + 组织测试 + 度量分析 + 输出报告。

设定典型任务的时候，我们要注意单个任务要有明确的开始和结束的动作、操作路径要明确、流程要简练；

在邀请目标用户的时候，我们只要邀请 5~8 名用户，就能够发现 80% 的问题；

组织测试开始之前，可以先进行一场预演；鼓励用户采用“发声法”完成测试，在测试过程中我们要做到客观记录，避免出现主观引导用户的操作；

度量分析包含了可用性度量和问题度量。对于 B 端产品来讲，可用性度量可以通过观察任务时长分布图得到一些更深入的结论，问题度量的一个关键点是辨别真问题和假问题；

最后，输出报告通常会包含两份资料：可用性测试报告和问题优化汇总表。可用性测试报告是对本次可用性测试的一个概括说明，问题优化汇总表是督促问题优化落地的一个工具。

只要任务设计清晰简练，在每个步骤中把握测试的原则，目标聚焦，即使一个人也可以做好用户可用性测试。

课后题

针对你现在服务的产品，如果你自己来安排一次可用性测试，你该怎样设定任务呢？在资源有限的情况下，你该怎样选择指标呢？

记得在留言区和我讨论、交流你的想法，每一次思考都会成为你进步的基石。

如果你喜欢今天的内容，也欢迎你把这一讲分享给你的朋友。

感谢你的阅读，我们下一讲再见。

提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 07 | 如何让你的设计更有温度？

下一篇 09 | 为什么你的设计方案总是通不过？

精选留言 (1)

写留言

石云升
2021-01-25

对小公司来说有点繁琐了。但这个方法确实可以发现那些设计过程中忽略的问题。而且是可以确定性的拿到问题报告。
有一个相对简单的找到问题的方法分享给到大家，那就是把注意力放在那些给你反馈问题的客户身上，千万不要让客服教他们解决问题，应该先记录问题，然后在帮助用户解决问题。对于这些记录的问题，需要针对性的做分析，有一部分是伪需求，这样我们也能得...
展开

作者回复: 棒！！

可用性这件事情不分公司大小，是看工作场景中是否需要，而且也可以根据人力和时间情况，在任务设定和邀请人员的数量上进行适当调整。可用性测试的目的是帮助我们发现产品设计相关问题的有效方式之一。

当然了，不管是什么产品，如果能够拿到来自于客户的一手反馈是最幸福的事情了。客服、应用市场、相关论坛、贴吧、知乎等都是了解客户意见的有效渠道。

1